The Development and Assessment of a Daily Rainfall Disaggregation Model for South Africa

D. M. Knoesen and J. C. Smithers

School of Bioresources Engineering and Environmental Hydrology
University of KwaZulu-Natal, P\Bag X01, Scottsville, South Africa, Email: beeh@ukzn.ac.za

Abstract

The temporal distribution of rainfall, *viz*. the distribution of rainfall intensity during a storm, is an important factor affecting the timing and magnitude of peak flow from a catchment and hence the flood-generating potential of rainfall events. It is also one of the primary inputs into hydrological models used for hydraulic design purposes. One method of accounting for the temporal distribution of rainfall is to disaggregate courser-scale data to a finer resolution, e.g. daily to hourly.

A daily to hourly disaggregation model developed in Australia has been modified and applied in South Africa. The primary part of the model is the distribution of R, which is the fraction of the daily total that occurs in the hour of maximum rainfall. A random number is used to sample from the distribution of R at the site of interest. The sample value of R determines the other 23 values, which then undergo a clustering procedure, together with the value for R, in order to best maintain the statistics for 2, 3, 6 and 12 hour durations. This clustered sequence can then be arranged into 1 of 24 possible temporal arrangements, depending on when the hour the maximum rainfall occurs. The structure of the model allows for the production of 480 different temporal distributions with variation between uniform and non-uniform rainfall.

The model was independently tested at 15 different locations in differing climatic regions in South Africa. At each location, observed hourly data were aggregated to yield daily values and were then disaggregated using the methodology. Results show that the model is able retain the daily total and the statistical characteristics of the hourly rainfall.

Keywords: *temporal rainfall disaggregation.*

1 Introduction

Engineers and hydrologists involved in the design of hydraulic structures, such as dams, bridges and culverts, need to accurately assess the frequency and magnitude of extreme hydrological events. Current techniques for design flood estimation in South Africa need to be updated, regional approaches need to be evaluated and new techniques have to be developed and applied (Smithers and Schulze, 2001). One input that would improve rainfall-runoff modelling, and hence the estimation of design floods, is to account for the temporal distribution of rainfall.

The temporal distribution of rainfall, *viz*. the distribution of rainfall intensity during a storm, is an important factor affecting the timing and magnitude of peak flow from a catchment and hence the flood-generating potential of rainfall events (Weddepohl, 1988). It is also one of the primary inputs into hydrological models used for the design of hydraulic structures. The temporal distribution of rainfall events may be influenced by many factors that need to be reflected in design temporal distributions. These factors include, *inter alia,* location, storm duration, storm depth, and season of storm occurrence (Hoang *et al.*, 1999).

The intensity distribution of a storm may be estimated by the use of a temporal distribution curve, which may be synthetically derived or obtained from observed hyetographs (Chow *et al.*, 1988). The use of temporal distributions is usually applicable to event-based models, such as the SCS-SA design flood estimation model (Schulze *et al.*, 1992). However, temporal distributions can also be used in rainfall disaggregation approaches (Boughton, 2000).

Rainfall disaggregation refers to producing high-resolution data that can be aggregated to give values equal to known courser-scale totals. The use of high-resolution rainfall data inherently accounts for the temporal distribution of rainfall intensity. This is because the incremental time-steps are small enough, i.e. hourly or sub-hourly, so as to represent different intensities. High-resolution rainfall data are often required as input into continuous simulation hydrological models. It is

important to note that disaggregation is not synonymous with downscaling. Downscaling aims at producing finer scale time-series with the required statistics, like disaggregation, but do not necessarily add up to any courser-scale totals (Koutsoyiannis, 2003). Downscaling is, in particular, used for hydrological applications of general circulation models.

Continuous simulation hydrological models are important tools when analysing complex hydrologic or hydraulic problems where issues need to be investigated at different timescales, for example in flood prediction and the modelling of water quality (Mikkelsen *et al.*, 1998). These models require detailed rainfall data, *viz*. hourly or sub-hourly. The advantage of such a time-series is that they reflect all relevant rainfall characteristics from peak intensities with short duration to variations in annual rainfall (Mikkelsen *et al.*, 1998). However, data are generally only widely available at more aggregated levels of the model time-step, such as daily. Koutsoyiannis and Onof (2001) note that in many countries, the number of raingauges providing hourly or sub-hourly resolution data is smaller than the number of daily gauges by about an order of magnitude. This situation reflects a general relative paucity of rainfall data for timescales of one hour or less, both in the number of gauges and length of the recorded series (Koutsoyiannis and Onof, 2001). This, too, is the case in South Africa where it is reported that there are 172 recording gauges with at least 10 years of breakpoint data (Smithers and Schulze, 2000a), compared to 1806 daily rainfall stations with at least 40 years of data (Smithers and Schulze, 2000b). The need for a model to disaggregate daily rainfall into a sequence of individual storms of finer timescale cannot be overemphasised (Gyasi-Agyei, 1999).

2 The Disaggregation Model

The daily to hourly disaggregation model used and modified in this study is based largely on the work done by Boughton (2000). The details of the model developed by Boughton (2000) are described in this section and changes to the methodology developed by Boughton (2000) are highlighted.

2.1 Structure of the Model

The model is comprised of 4 main parts:

a) The distribution of the fraction of the daily total, R, that occurs in the hour of maximum rainfall.
b) For each value of R there is an average set of values for the other 23 hourly fractions of the daily total.
c) Given the 24 fractions from above, the values are clustered to maintain the observed average highest 2-hour, 3-hour, 6-hour and 12-hour fractions.
d) These clusters are then arranged into random patterns so as to reproduce the possible variations in daily temporal patterns while retaining the abovementioned statistics.

2.2 Distribution of R

The primary part of the disaggregation model is the fraction, R, of the daily total that occurs in the hour of maximum rainfall. A value of R = 1.0 indicates that all of the rainfall on the day fell in a single hour. This is the upper limit of R and is the boundary of non-uniformity. Completely uniform rainfall throughout a day would yield R = 0.04167 (i.e. 1/24 of the daily total). This is the lower limit of R.

An example of a single day's rainfall in hourly increments at Raingauge N23 at Ntabamhlope in the KwaZulu-Natal (KZN) midlands is shown in Figure 1. The daily rainfall total was 84.4 mm and the hour in which the most rainfall fell contained 40.9 mm. This yields a ratio of R = 40.9/84.4 = 0.48 for the day.

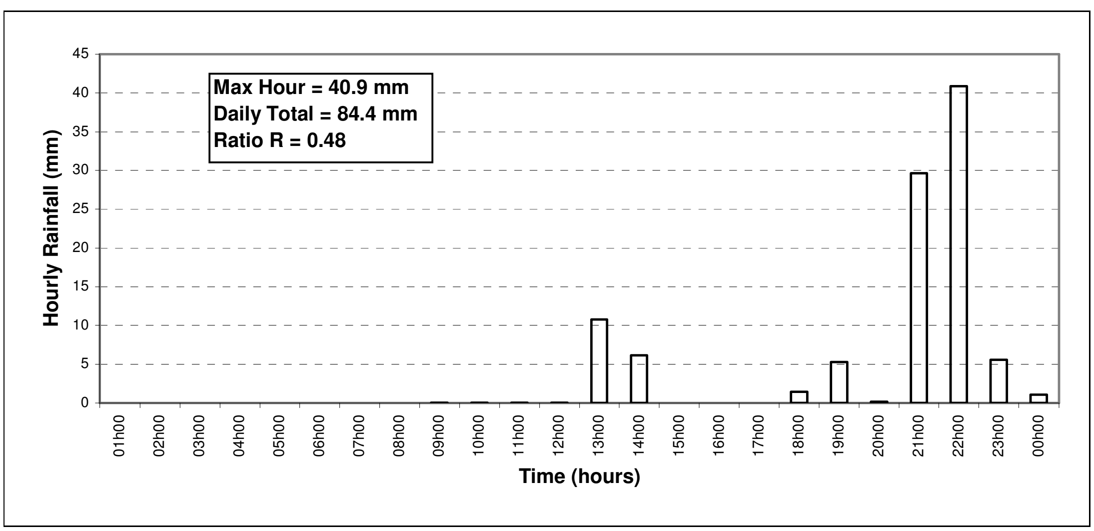

Figure 1. Example of a single day's hourly rainfall at Ntabamhlope

If the ratio R is determined for all days with 1 mm of rainfall or more, the distribution of R has a pattern that is a major characteristic of hourly rainfall at the site. The distribution of R for a particular site is created by extracting all the values of R at the site for days where the rainfall was greater than or equal to 1 mm, for the entire length of record. The computed R values are then collated into 20 ranges, which were used by Boughton (2000) and are shown in Table 1. The distribution of R thus shows the proportion of all values of R in each of the ranges.

Table 1. Ranges used when collating R values

No.	Range	No.	Range	No.	Range	No.	Range
1	0.0417-0.075	6	0.275-0.325	11	0.525-0.575	16	0.775-0.825
2	0.075-0.125	7	0.325-0.375	12	0.575-0.625	17	0.825-0.875
3	0.125-0.175	8	0.375-0.425	13	0.625-0.675	18	0.875-0.925
4	0.175-0.225	9	0.425-0.475	14	0.675-0.725	19	0.925-0.975
5	0.225-0.275	10	0.475-0.525	15	0.725-0.775	20	0.975-1.000

The distributions of R for two sites in differing climates are shown in Figure 2. Jonkershoek (Station Jnk19a), in the Western Cape, is located in a winter rainfall region whereas Ntabamhlope (Station N23),in the KZN midlands, is located in a summer rainfall region.

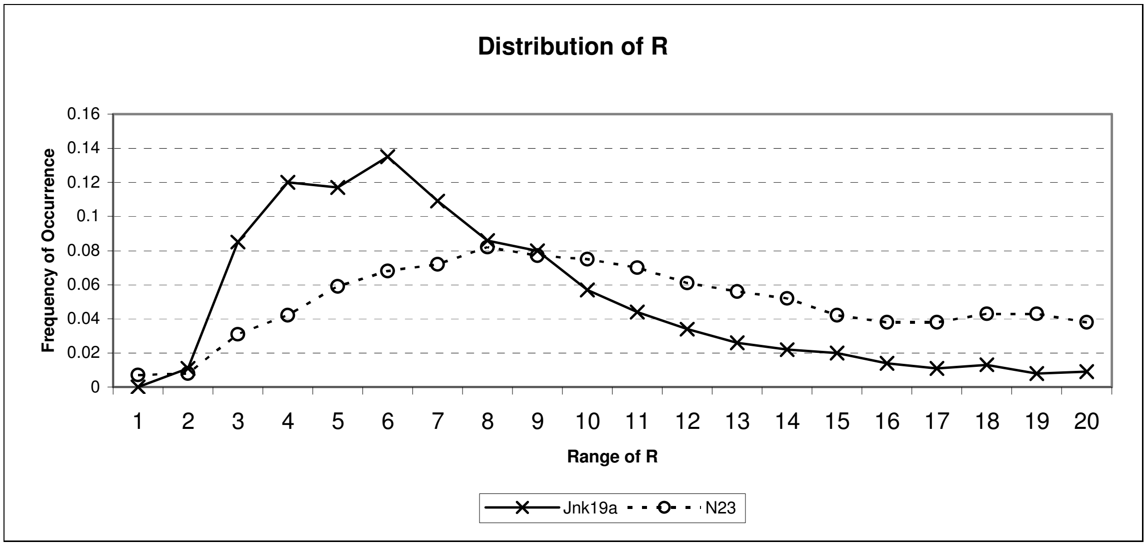

Figure 2. Frequency distributions of R at Stations Jnk19a (in the W. Cape) and N23 (in KZN).

From Figure 2 it is evident that the majority of the days at Jonkershoek fall into Range 6 in Table 1 and have many low values of R (mean R = 0.385) indicating that there is a tendency for more uniform rainfall. The distribution of R at Ntabamhlope (mean R = 0.537) shows a larger proportion of the days having higher values for R. This indicates that at Ntabamhlope large portions of the daily rain fall in a single hour, which is typical of the convective storms in the summer rainfall region.

2.3 Calculating the Other 23 Hourly Fractions

If R = 1.0 then all of the rainfall fell in a single hour, hence the other 23 hourly fractions must be 0. If R = 0.04167 then the other 23 hourly fractions must equal 0.04167. If, however, R is slightly less than 1.0 it is probable that the rest of the day's rainfall fell in 1 or 2 other hours, resulting in the remaining 21 or 22 hours having zero rainfall. Conversely, if R is slightly greater than 0.04167 then the other 23 values will probably be slightly less than, but close to, 0.04167. This is important to note as it indicates that the value of R has a strong influence in determining the other 23 hourly fractions of rainfall.

In order to determine the other 23 hourly fractions, the 24 hourly fractions for every day on record were ranked in order of magnitude, with R being the largest value on each day. This was done for 157 stations. These ranked series, from all 157 sites, were then pooled together and the values for each rank were averaged in each of the 20 ranges of R shown in Table 1. This resulted in 20 averaged ranked series of hourly fractions, one for each range of R.

Once all 24 hourly averaged fractions have been determined for each range of R they can be used to create daily temporal patterns of rainfall. The following two sections contain a description of how these 24 hourly fractions are arranged to recreate possible realisations of the temporal distribution of daily rainfall.

2.4 Clustering of Hourly Rainfalls

In order to cluster the 24 hourly fractions, the data from all stations were again processed to calculate the highest 2-hour fraction of the daily total, the highest 3-hour fraction, the highest 6-hour fraction and the highest 12-hour fraction on each day. As for the ranked series, all of these fractions were then averaged within the range of R in which they occurred. This

resulted in an average 2-hour fraction, 3-hour fraction, 6-hour fraction and 12-hour fraction of the daily total for each of the 20 ranges of R.

Using the abovementioned ranked sequences, a computer program was used to check the sum of the first value in the ranked series with each of the other 23 hourly fractions in order to find which of the 23 values gave the best match to the average 2-hour fraction for the respective range of R. After identifying the value in the series, which when added to the first value best approximated the average observed highest 2-hour fraction, the program then checks the remaining 22 hourly values to find which value should accompany the 2-hour fraction to approximate the average highest 3-hour fraction. The program then searches the remaining 21 values for a combination of 3 values to form the average highest 6-hour fraction, and then searches for the next combination of 6 values to form the average highest 12-hour fraction. Performing this for each range of R resulted in 20 clustered sequences. The next step was to arrange these clustered sequences into temporal patterns.

2.5 Daily Temporal Patterns of Hourly Rainfalls

Schmidt and Schulze (1987) derived four design rainfall temporal distributions to be used for different regions in South Africa. This suggests that a single distribution can be used to represent the temporal distribution of rainfall for a particular region. This, however, is not realistic and analysis of the rainfall data shows that there are several temporal patterns ranging from nearly uniform rainfall to highly variable rainfall. Furthermore, the peak intensity can occur during any hour of the day, adding to the variability of temporal rainfall patterns. In order to attempt to account for the variability of temporal patterns of rainfall, several temporal distributions should be employed.

The hour of day when the highest intensity rainfall occurred was determined for each station. The results show a definitive distribution for the timing of peak rainfall occurrence at a particular location. As shown in Figure 3 for Station Jnk19a, the hour of maximum rainfall has a somewhat uniform distribution, indicating that the hour of maximum rainfall has a reasonably equal probability of occurring in any hour on a particular day. Station N23, however, has a sinusoidal-like distribution with the majority of days having the peak rainfall falling during the late afternoon and evening.

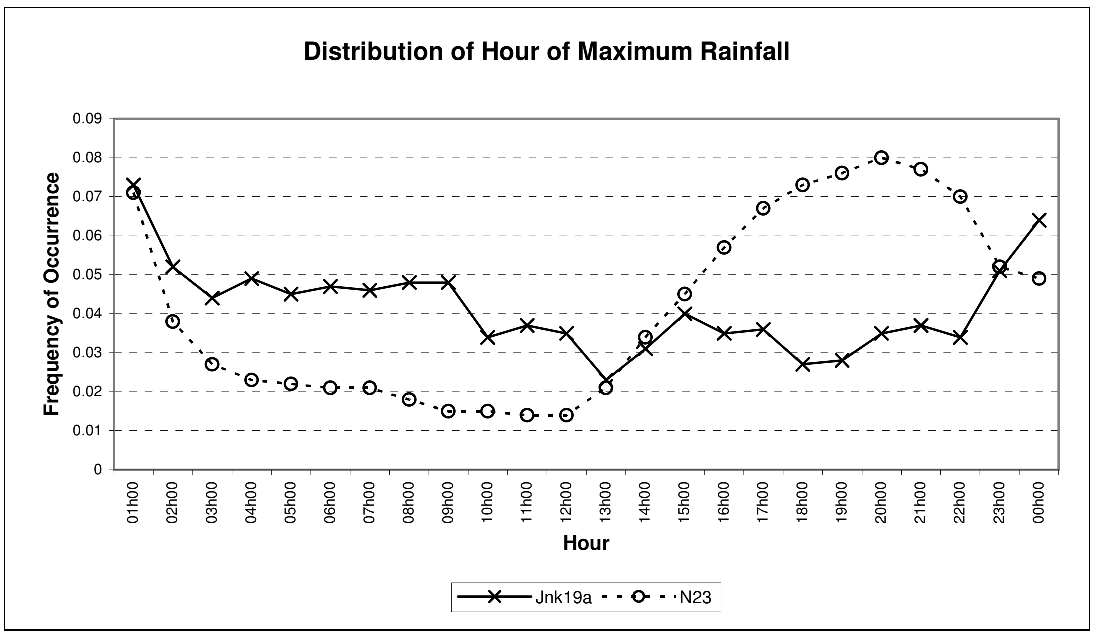

Figure 3. Frequency distributions of the hour of maximum rainfall at Jonkershoek (in the W. Cape) and Ntabamhlope (in KZN)

When applying the disaggregation method, a random number is used to select the hour of maximum rainfall from the distribution of the hour of maximum rain for the site of interest. This differs from the work done by Boughton (2000) as in that study no distinct distribution was found for the time of maximum rainfall and hence the hour of maximum rainfall was selected at random.

Using the clustered sequences established above, and then accounting for all permutations when the hour of maximum rainfall can occur, 24 arrangements of the clustered sequences can be created. The combination of these 24 arrangements with the 20 possible ranges of R results in 480 different temporal patterns, as opposed to one averaged distribution currently commonly used in South Africa. These range from uniform to non-uniform with the possibility of the hour of maximum rainfall occurring in any hour of the day. Figure 4 contains a sample of the different temporal distributions that the disaggregation model produces.

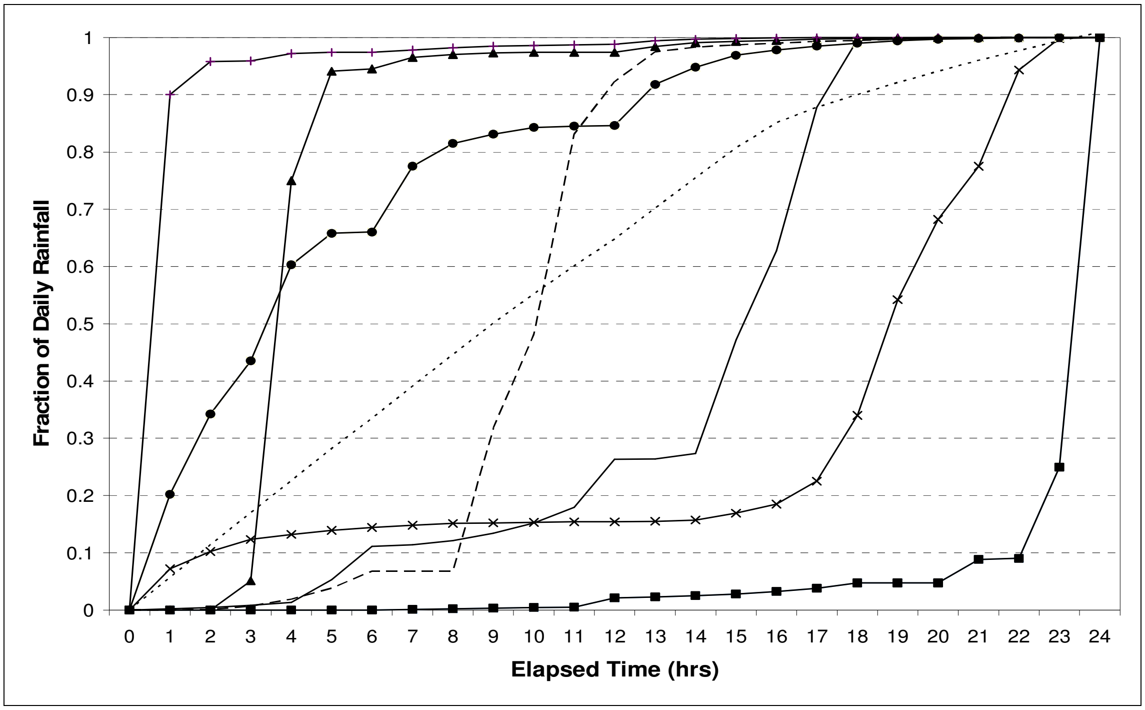

Figure 4. Samples of the different temporal distributions that are generated by the disaggregation model

3 Model Testing and Results

In order to quantify the simulated performance of the disaggregation model, a similar approach to that used by Smithers and Schulze (2000a) was employed. Moments and other event characteristics computed from the disaggregated rainfall series are compared to the equivalent values computed from the observed data in Section 3.1. Similarly, design rainfall depths computed from the disaggregated rainfall series are compared to the equivalent values computed from the observed data in Section 3.2. Both measures of quantifying the model performance were carried out at 15 test stations, details of which are given in Knoesen (2005). The data from all 15 test stations were excluded from model development.

3.1 Moments and Statistics

The two random processes that occur within the disaggregation model, *viz.*, the selection of the value of R and the timing of the hour of maximum rainfall, introduce stochastic variability in the output. At each of the selected test stations the stochastic variability was simulated by generating one hundred disaggregated series. A frequency analysis was performed on the 100 sets of disaggregated values for all statistics and durations considered. High-Low bar graphs depicting the observed moments and the 25th and 75th non-exceedance percentiles of the 100 sets of disaggregated values are used to graphically depict the performance of the model. The poorest results were achieved at Station 0435019, located at Ottosdal, while the best results were achieved at Station Sacfs, located at Umhlanga. The results from disaggregating the 24-hour data at Stations 0435019 and Sacfs are shown in Figures 5 and 6 respectively.

It can be seen in Figure 5 and Figure 6 that the disaggregation model performs equally well in simulating the mean, standard deviation and skewness at Stations 0435019 Sacfs for the range of durations considered. It was expected that the mean rainfall for all levels of aggregation should be simulated extremely well owing to the method of disaggregation. The model tends to be less capable of simulating certain event characteristics and statistics such as event duration and dry probability. This is a weakness in the disaggregation model and suggests that more work needs to be done on refining the sequencing of the hourly rainfalls. The distinguishing factors between the best and worst simulations are the lag 1 – 10 autocorrelations, and is shown by Knoesen (2005) to be directly related to the quality of the data at the respective sites.

Furthermore, it is evident from Figures 5 and 6 that the dry probabilities, and hence the event durations, are better simulated at Station 0435019 than at Station Sacfs. It is postulated that this may be attributed to the distribution of R at Station 0435019, which has a mean value of R (R_{mean}) of 0.607, whereas Station Sacfs has an R_{mean} of 0.499. This appears to be a shortcoming of using a single distribution of R to represent all rainfall depths. Although the disaggregated data will have the correct overall distribution of R, it is likely that the tails of the rainfall distribution will not receive the correct values of R.

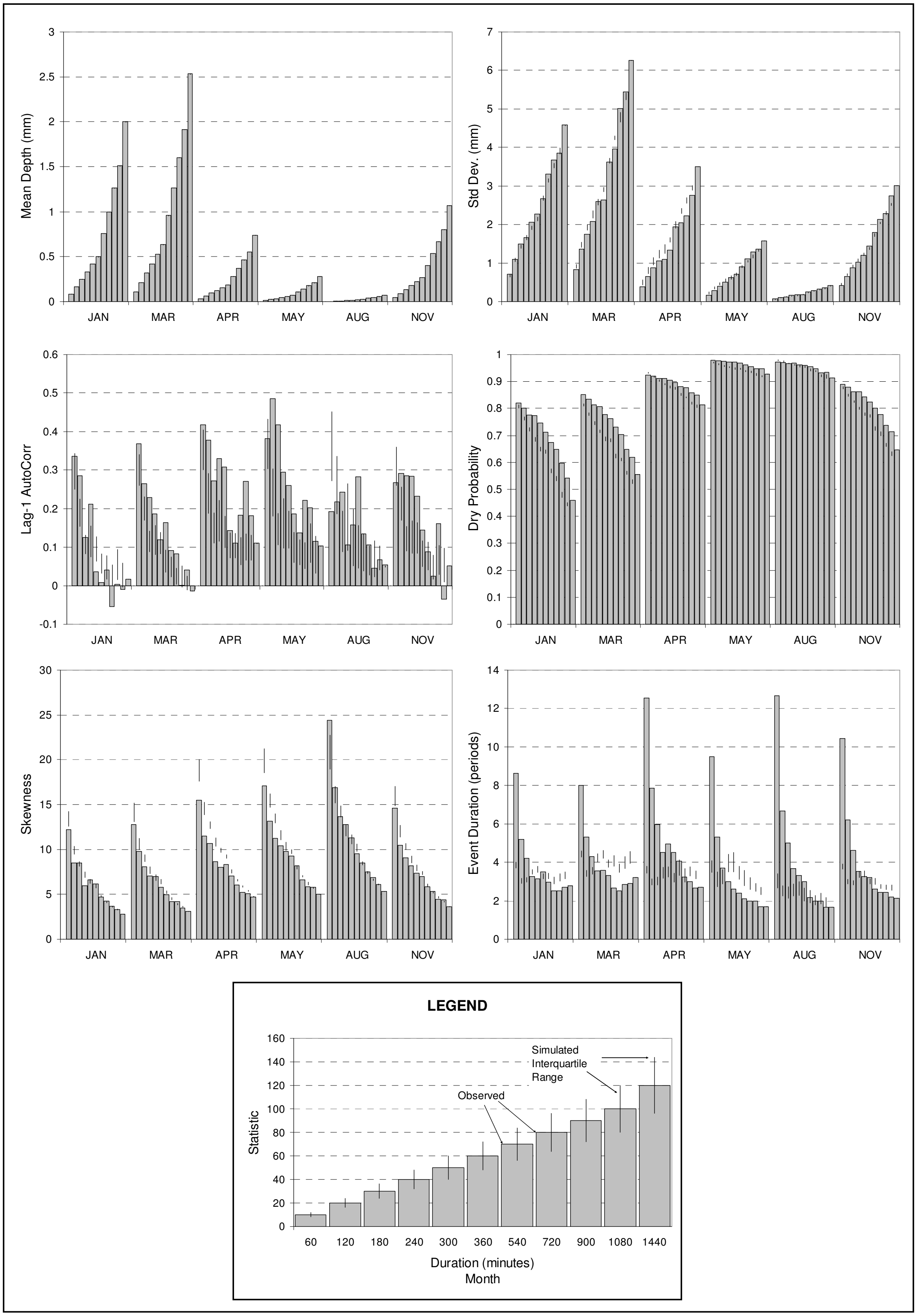

Figure 5. Simulated performance of the disaggregation model at Station 0435019

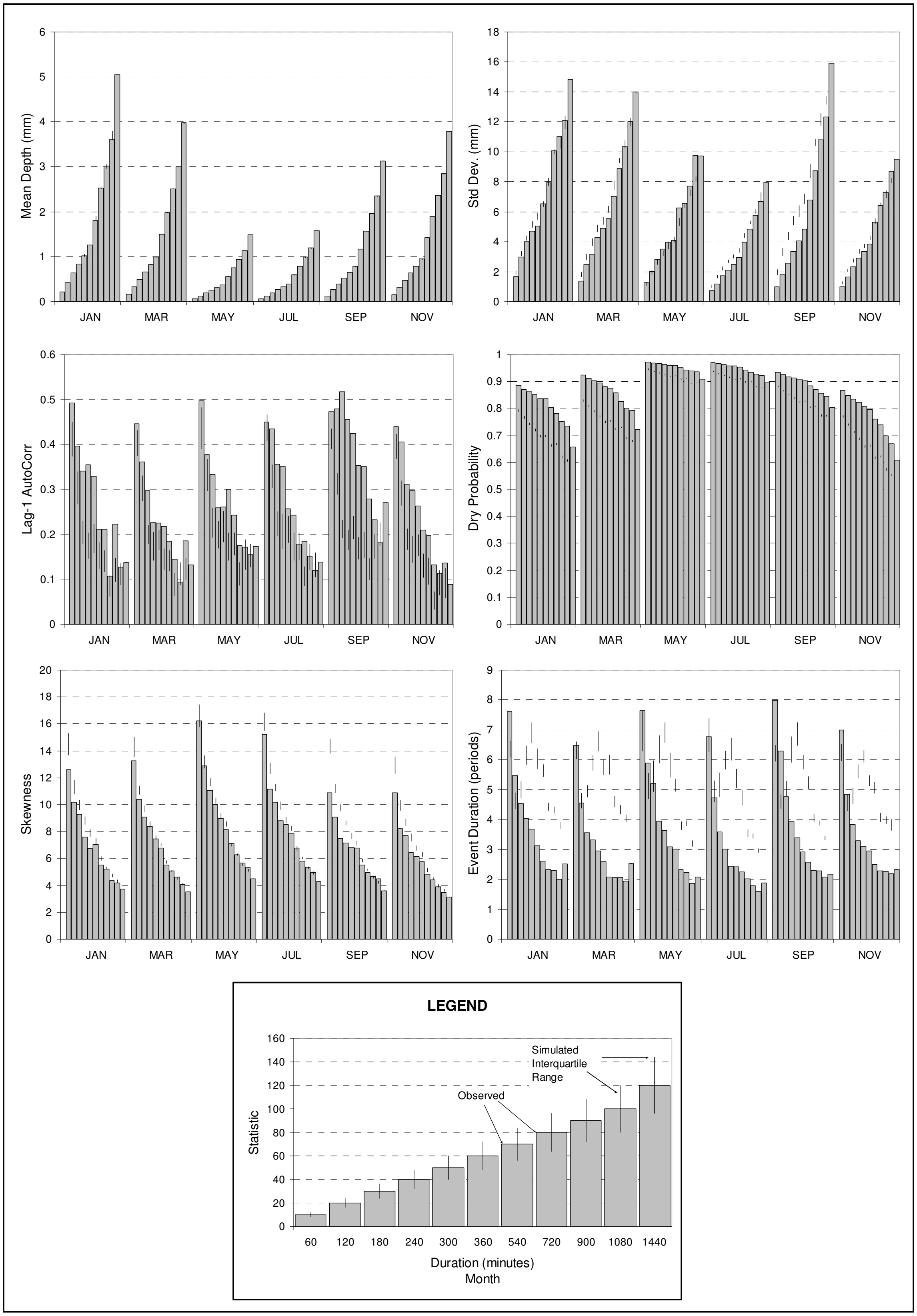

Figure 6. Simulated performance of the disaggregation model at Station Sacfs

3.2 Extreme Rainfall Events

Similar to the procedures used by Smithers and Schulze (2000a), design rainfall depths were calculated using the General Extreme Value (GEV) distribution fitted to the Annual Maximum Series (AMS) by L-moments, for the observed data and for each of the 100 disaggregated series generated from the disaggregation model. Design values for the 2, 5, 10, 20, 50, and 100-year return periods were computed for durations of 1, 2, 3, 4, 6, 8, 10, 12, 16, 20 and 24-hour. For each duration and return period, a frequency analysis was performed on the 100 values computed from the disaggregated rainfall series generated by the disaggregation model. High-Low bar graphs depicting the observed design rainfall computed from the observed data and the 25th and 75th non-exceedance percentiles of the design rainfall computed from the 100 disaggregated rainfall series were used to evaluate the performance of the model. Examples of model performance, with respect to design rainfall estimation, are shown in Figure 7, which depict the worst (Station 0028748) and best (Station 0474680) simulations.

The poor performance observed when estimating design floods at Station 0028748 seems to be related to the distribution R, i.e. the station that displayed the best results has an R_{mean} = 0.60, whereas the station with the worst results, Station 0028748, has an R_{mean} value = 0.45. After analysing all the test stations it was found that the stations with the highest R_{mean} values gave the best results. This is because on those days when smaller rainfall events (± 1 mm) occurred it is likely that the all the day's rainfall fell within a few hours, thus unduly influencing the distribution R. Although this will affect all the stations used, it appears that the error is exacerbated for those stations with lower mean R values. It is postulated that the use of different distributions of R to represent rainfalls of differing magnitudes will improve the performance of the rainfall disaggregation model, particularly in the estimation of design rainfall.

3.3 Regionalising the Methodology

In order to apply the methodology at sites where no short-duration data are available, it is necessary to regionalise the methodology. Assessing the model performance according to the moments and statistics of the disaggregated series, as well as the design rainfall computed from the disaggregated series, results obtained by Knoesen (2005) show that applying the model using regionalised input gives very similar results to those obtained when at-site short duration data are available. For details regarding the approach used to regionalise the methodology, the reader is referred to Knoesen (2005).

4. Discussion

The rainfall disaggregation model developed by Boughton (2000) was intended to be used only for design flood purposes, and hence only focused on disaggregating larger daily rainfalls (>15mm). In order to achieve the objectives of this study, the methodology was modified. All days for which the aggregated 24-hour rainfall total was greater than or equal to 1 mm were used to develop the disaggregation model, which is used to disaggregate all non-zero 24-hour rainfalls.

Further modifications were made to the methodology regarding the distribution of the time when the hour of maximum rainfall occurred. It was evident that the rainfall data at different stations displayed different distributions for the hour of maximum rainfall. The distribution of the hour of maximum rainfall at each station was computed, and random sampling from the respective distributions is performed. This differs from the original Boughton (2000b) model where the hour of maximum rainfall was determined by random sampling from a uniform distribution.

The resulting model is capable of producing 480 different temporal patterns with ranging levels of uniformity. The distribution of R and the distribution of the hour of maximum rainfall for each station determine which of the 480 possible temporal patterns to select for a particular day's rainfall.

Two measures were employed in order to quantify the performance of the disaggregation model. Firstly, moments and other event characteristics were computed from the disaggregated data and compared to the equivalent values computed from the observed data. Secondly, design rainfall depths were computed from the disaggregated data and compared to the equivalent values computed from the observed data.

Owing to the stochastic nature of the disaggregation models, 100 disaggregated series were generated for each independent test location and frequency analyses were performed. High-Low bar graphs depicting the observed moments and the 25th and 75th non-exceedance percentiles of the 100 sets of disaggregated data were used to graphically depict the adequacy of the model.

The results from the model where at-site short duration data are available indicate that the model is able to produce synthetic hourly data which resemble the general distribution of the observed hourly data for a particular site. However, the results also indicate that the model is less capable of simulating some of the statistics considered i.e. the probability of dry periods and design rainfalls for selected return periods.

The relatively poor results obtained for the various lag autocorrelations from the disaggregated rainfall is postulated to be the result of the way that the hourly rainfalls are sequenced. Although the lag autocorrelations were better simulated for longer and better quality data, it is suggested that a different method of sequencing the disaggregated hourly rainfalls may improve the model in this respect.

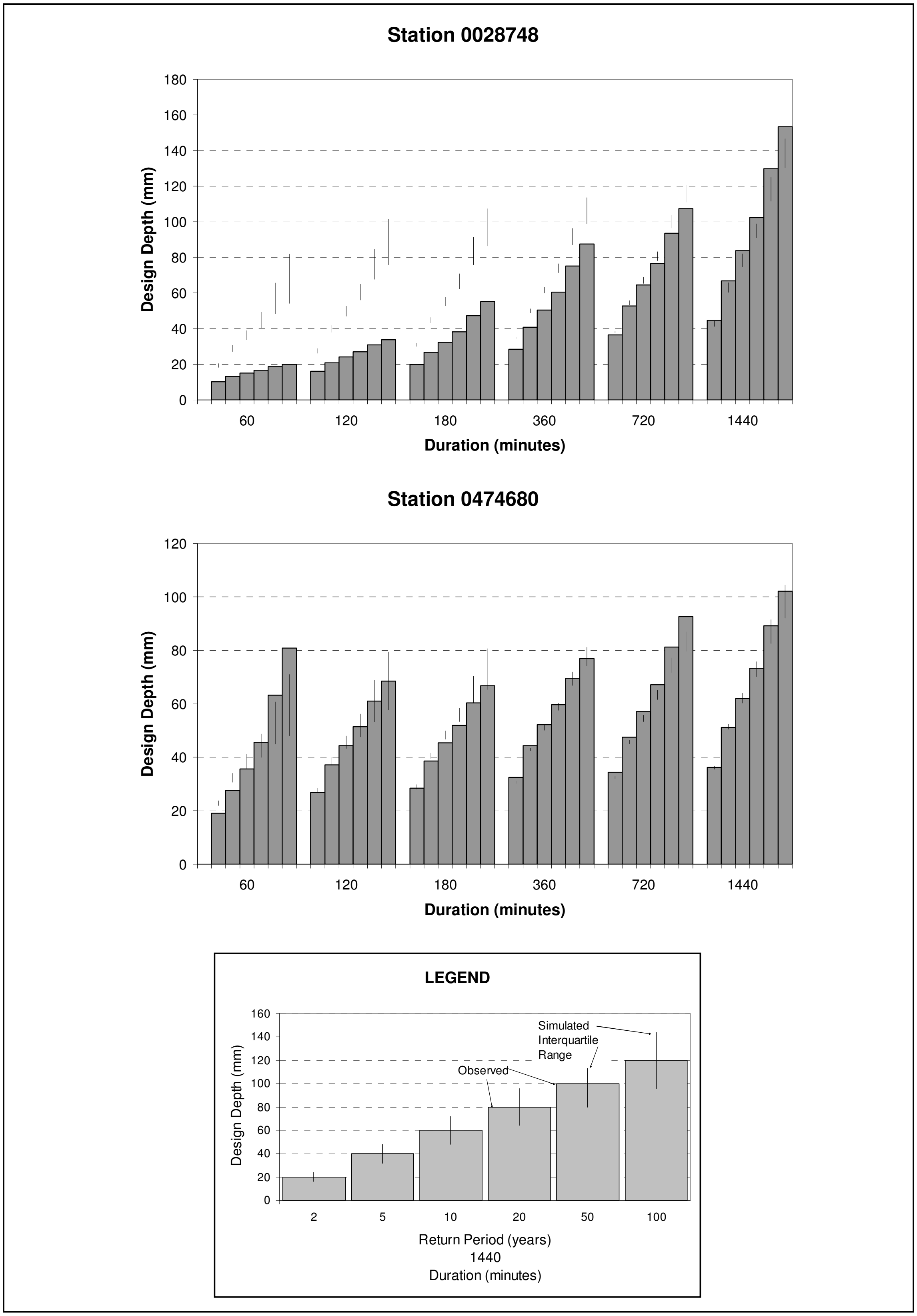

Figure 7. Design rainfall estimated using disaggregated data for Stations 0028748, at George, and 0474680, at Carletonville

The methodology was regionalised, which is a further modification to the methodology employed by Boughton (2000), and the results obtained when using the regionalised input to the model are similar to those obtained when at-site information is available, for both statistical and extreme value measures of model performance. This is a positive result as it implies that the model can be used to reasonably disaggregate daily rainfall at locations in South Africa where no short duration data are available.

5. Conclusions

From this study it can be concluded that:

- The temporal distribution of rainfall is an important factor affecting the timing and magnitude of peak discharge from a catchment.
- A methodology for the disaggregation of daily rainfall to produce hourly increments which aggregate to equal the observed daily values has been identified and applied in South Africa.
- The methodology is shown to reproduce the general distribution of rainfall relatively well, both when observed short duration data are available as well as in the absence of such information.
- Refinements need to be made to the methodology in order to improve sequencing and the correlation structure of the disaggregated rainfall.

6. Recommendations for Future Research

It is postulated that the weakness of the model in simulating both extremes of the rainfall spectrum (dry probabilities and design rainfall) is a result of the use of a single distribution of R to represent an entire range of magnitudes within a rainfall record. It is recommended that, for a particular rainfall record, the data be collated according to the daily rainfall total, using pre-determined ranges, and a distribution of R be calculated for each of these ranges. It is further recommended that more research be done on how to sequence the disaggregated hourly rainfalls, in order to improve the simulation of the structure of the rainfall, as measured by the lag autocorrelations, number of events and event durations.

It is further recommended that the disaggregation model be linked with a reliable daily rainfall generator. This would facilitate the generation of long sequences of hourly data for any location in South Africa which could be used for modelling of water resources and design flood estimation.

References

Boughton, W. 2000. A model for disaggregating daily to hourly rainfalls for design flood estimation. Cooperative Research Centre for Catchment Hydrology, Monash University, Clayton, Victoria, Australia. Report 00/15, 36 pp.

Chow, V.T., Maidment, D.R. and Mays, L.W. 1988. Applied hydrology. McGraw-Hill, New York, USA.

Gyasi-Agyei, Y. 1999. Identification of regional parameters of a stochastic model for rainfall disaggregation. Journal of Hydrology, 223: 148-163.

Hoang, T.M.T., Rahman, A., Weinmann, P.E., Laurenson, E.M. and Nathan, R.J. 1999. Joint probability description of design rainfall. Conference on Water Resources and Environmental Research, Brisbane, Australia.

Knoesen, D.M. 2005. The Development and Assessment of Techniques for Daily Rainfall Disaggregation in South Africa. Unpublished M.Sc. Thesis, University of KwaZulu-Natal, Pietermaritzburg, RSA, 88 pp.

Koutsoyiannis, D. 2003. Rainfall disaggregation methods: Theory and applications. Workshop on Statistical and Mathematical Methods for Hydrological Analysis, Rome, Italy.

Koutsoyiannis, D. and Onof, C. 2001. Rainfall disaggregation using adjusting procedures on a Poisson cluster model. Journal of Hydrology, 246: 109-122.

Mikkelsen, P.S., Madsen, H., Arnbjerg Nielsen, K., Jorgensen, H.K., Rosbjerg, D. and Harremoes, P. 1998. A rationale for using local and regional point rainfall data for design and analysis of urban storm drainage systems. Water Science and Technology, 37(11): 7-14.

Schmidt, E.J. and Schulze, R.E. 1987. SCS-based design runoff. Department of Agricultural Engineering, University of Natal, Pietermaritzburg, RSA. ACRU Report No. 24, 164 pp.

Schulze, R.E., Schmidt, E.J. and Smithers, J.C. 1992. PC-Based SCS design flood estimates for small catchments in southern Africa. Department of Agricultural Engineering, University of Natal, Pietermaritzburg, RSA. ACRU Report 40, 47 pp.

Smithers, J.C. and Schulze, R.E. 2000a. Development and evaluation of techniques for estimating short duration design rainfall in South Africa. Water Research Commission, Pretoria, RSA. WRC Report No. 681/1/00, 356 pp.

Smithers, J.C. and Schulze, R.E. 2000b. Long duration design rainfall estimates for South Africa. Water Research Commission, Pretoria, RSA. WRC Report No. 811/1/00, 69 pp.

Smithers, J.C. and Schulze, R.E. 2001. Design runoff estimation: A review with references to practices in South Africa. Tenth South African National Hydrology Symposium, SANCIAHS, CSIR. Pietermaritzburg, RSA.

Weddepohl, J.P. 1988. Design rainfall distributions for Southern Africa. Unpublished M.Sc. Thesis, Department of Agricultural Engineering, University of Natal, Pietermaritzburg, RSA, 162 pp.